

## Hamline University DigitalCommons@Hamline

---

School of Education Student Capstone Theses and  
Dissertations

School of Education

---

Fall 12-12-2016

# MEASURING WHAT STUDENTS KNOW: ENSURING MATH WORD PROBLEMS ON SCHOOL WRITTEN ASSESSMENTS ACCURATELY DISPLAY ENGLISH LANGUAGE LEARNERS' UNDERSTANDING OF MATH SKILLS

Lori Magstadt  
*Hamline University*

Follow this and additional works at: [https://digitalcommons.hamline.edu/hse\\_all](https://digitalcommons.hamline.edu/hse_all)

 Part of the [Education Commons](#)

---

### Recommended Citation

Magstadt, Lori, "MEASURING WHAT STUDENTS KNOW: ENSURING MATH WORD PROBLEMS ON SCHOOL WRITTEN ASSESSMENTS ACCURATELY DISPLAY ENGLISH LANGUAGE LEARNERS' UNDERSTANDING OF MATH SKILLS" (2016). *School of Education Student Capstone Theses and Dissertations*. 4251.  
[https://digitalcommons.hamline.edu/hse\\_all/4251](https://digitalcommons.hamline.edu/hse_all/4251)

This Thesis is brought to you for free and open access by the School of Education at DigitalCommons@Hamline. It has been accepted for inclusion in School of Education Student Capstone Theses and Dissertations by an authorized administrator of DigitalCommons@Hamline. For more information, please contact [digitalcommons@hamline.edu](mailto:digitalcommons@hamline.edu), [lterveer01@hamline.edu](mailto:lterveer01@hamline.edu).

MEASURING WHAT STUDENTS KNOW: ENSURING MATH WORD PROBLEMS  
ON SCHOOL WRITTEN ASSESSMENTS ACCURATELY DISPLAY ENGLISH  
LANGUAGE LEARNERS' UNDERSTANDING OF MATH SKILLS

by

Lori Magstadt

A capstone submitted in partial fulfillment of the requirements for the degree of  
Master of Arts in English as a Second Language

Hamline University

Saint Paul, Minnesota

December 2016

Committee:

Feride Erku, Primary Advisor

Angela Mansfield, Secondary Advisor

Christine Hubred, Peer Reviewer

## TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION .....	1
Assessment Process .....	3
Challenges of the Assessment.....	4
Role and Background of the Researcher.....	5
Research Question .....	9
Research and Method.....	9
Mapping out the Capstone .....	10
CHAPTER TWO: LITERATURE REVIEW .....	12
A Brief History of Assessment .....	12
Current Math Standardized Testing in Minnesota .....	19
Challenges in Assessing ELs in Math.....	23
What is Universal Design?.....	27
School Response to Standardized Testing: Interim Assessments .....	34
The Gap in Current Research and Research Questions .....	37
Summary .....	37
CHAPTER THREE: METHODS .....	39
Overview of Chapter.....	39
Research Paradigm .....	40
Data Collection .....	42
Participants.....	42
Setting .....	42
Materials .....	43
Data Collection Technique 1: Selecting the 60 Math Word Problems .....	44
Procedure .....	48
Method of Analysis.....	53
Verification of Data .....	55

Ethics .....	55
Summary .....	55
CHAPTER FOUR: RESULTS .....	57
Results Based on Grade Level, Standard and DOK.....	64
Conclusion .....	66
CHAPTER FIVE: CONCLUSIONS .....	68
Major Findings.....	68
Limitations.....	72
Implications and Further Steps .....	73
Further Research .....	74
APPENDIX A.....	75
References.....	76

## LIST OF TABLES

Table 1: Webb’s Depth of Knowledge Key Actions (Webb, 2002) .....	3
Table 2: Range of Items per Strand for Mathematics MCA-III Grade 3-8 .....	20
Table 3: Cognitive Level Target Minimum Distribution.....	20
Table 4: Levels 1 and 2 Depth of Knowledge Standard 3.1.2.1 .....	21
Table 5 : Schools with Access to the Question Item Word Bank.....	42
Table 6: 4 <sup>th</sup> Grade Math Item Sample.....	43
Table 7: Criteria for Questions Selected.....	48
Table 8: DOK and Standard Alignment.....	49
Table 9: 4 <sup>th</sup> Grade Pilot Study Results.....	51
Table 1	
0: Factorial Design.....	54
Table 11: Passive Voice Items .....	58
Table 12: Proper Nouns .....	59
Table 13: Math Specific Vocabulary .....	62
Table 14: Universal Context .....	63
Table 15: Number of Modification Violations .....	64



## CHAPTER ONE: INTRODUCTION

As an educational consultant, I am constantly asked to support teachers in assessing their own instruction. I am currently in my third year of working with a group of charter schools to improve student learning. Before this, I worked at a charter school in the Southwest Metro of Minneapolis as the Academic Coordinator for five years. During that time, I implemented an initiative called Data Driven Instruction (DDI), where teachers develop quarterly assessments based solely on what was taught to ensure students are learning. In the last three years, I have worked with eleven additional schools to develop assessments, imbed formative assessment tools into lessons, and coach teachers on ways to change future instruction based on the results of assessments. The schools I work with range from small rural schools to medium-sized intercity schools. Similarly, the schools' populations are diverse. Five of the schools are at least 50% EL and at least 85% Free and Reduced Lunch. The others hold a smaller percent of EL, but only one does not have any enrolled.

My experience in so many different settings has shown me that no matter the demographics of the students or other factors of the environment, a clear idea of what will be taught and to what level of rigor must be in place before instruction occurs. Standards may describe skills that students need to learn, but I have found that teachers have a wide variety of understanding of these standards based on their experience and previous education. The most basic way to come to a common understanding amongst all teachers is to create actual questions that all students should be able to answer if they

have learned that skill. The challenge then becomes ensuring these questions are free of complex sentence structures and culturally relevant content to ensure that only the skill taught is being assessed. Teachers do not have the time or training to ensure math word problems are linguistically ethical like the developers of standardized assessments.

Therefore, I want to see how consistently the main linguistic principles used to develop standardized math word problems are applied to teacher written interim assessments. I will then use this information to redesign my professional development trainings around math word problems to include my findings.

In this chapter, I will expound on the development of the assessment process and the selection of the specific questions used to determine if a student has learned the targeted math skills taught. For example, if a teacher has just finished a unit on double digit addition and subtraction and their use in everyday life, the word problem used to assess this needs to use a situation that all students experience and vocabulary appropriate to the situation selected. I will then explain the specific question raised in the language development of math word problems to ensure that these questions are equitable for all students, no matter their level of English development. Next I will layout my role as a researcher and my background. Then I will present the two questions I hope to answer and the method I used to go about answering these questions. Lastly, I will map out the rest of the paper.



## Assessment Process

With an increase of standards based instruction and assessment, teachers continually rely on assessment results to define what skills students have mastered and on what skills students need continued practice and instruction (Bambrick-Sontoyo, 2010). I am currently working with a group of charter schools that have adopted DDI strategies, where all students take a summative assessment crafted by teachers and teacher leaders in order to ensure students are learning what has been taught. While not everything that makes students successful after school can be measured on an assessment, the hope is that these assessments give a clear picture of what students have learned during instruction and therefore are crafted to reflect the language used during class. Along with skills based on state standards, math assessments contain questions addressing three different levels based on Webb's Depth of Knowledge, or simply DOK (2002). This ensures multiple levels of depth, from basic recall to application of skills in multiple situations. Table 1 shows a greater description of these levels.

**Table 1: Webb's Depth of Knowledge Key Actions (Webb, 2002)**

<b>Depth of Knowledge</b>	<b>Key Actions</b>
DOK 1 ( <i>Recall</i> )	List, Tell, Define, Label, Identify, Name, State, Write, Locate, Find, Match, Measure, Repeat
DOK 2 ( <i>Skill/Concept</i> )	Estimate, Compare, Organize, Interpret, Modify, Predict, Cause/Effect, Summarize, Graph, Classify
DOK 3 ( <i>Strategic Thinking</i> )	Critique, Formulate, Hypothesize, Construct, Revise, Investigate, Differentiate, Compare
DOK 4 ( <i>Extended Thinking</i> )	Design, Connect, Synthesize, Apply, Critique, Analyze, Create, Prove, Support

To put it simply, each math question assesses at least two aspects for all students: 1) If they have learned the skill taught and 2) to what level they understand this skill. Math word problems have a third component that is assessed: 3) language processing. I will go into greater depth of these three aspects in Chapter 2.

### Challenges of the Assessment

Challenges of crafting a comprehensive assessment that checks exactly what the student did or did not learn after it is taught include time to create the assessment; clear understanding of who on the teaching team is creating the assessment; and that the teaching team has adequate training to write or select assessment questions that reflect what was taught (Bambrick-Sontoyo, 2010). From my experience, teachers often struggle to balance these measurable, academic goals with life skills also necessary for students to acquire while in school. Communication skills, executive functioning, and grit are only a few examples of skills that all students need to be successful in college and the work place that cannot be measured on a paper-pencil or online test. Teachers are faced daily with the challenge to balance fostering social skills as mentioned above with state academic expectations. Additionally, teachers in the standard elementary classroom are teaching on average four different subjects each day. Between lesson planning and classroom management, the task of creating assessments can be overwhelming. Since the majority of BA educational degrees do not require an assessment development course, teachers are often tasked to create assessments with little training or professional

development. Due to these complications, my experience shows that teachers approach assessment writing for lessons they are about to teach with caution and sometimes fear. However, time and support in crafting assessments can decrease anxiety and teachers can find that a clear, defined path can make lesson development easier and more consistent (Bambrick-Sontoyo, 2010).

Due to these challenges, I have created a professional development strategy that I have implemented in five different schools to support teachers for at least a year as they craft their own assessments. Before crafting the assessment, I meet with the team of teachers and support them as they determine how many questions each standard or skill should represent and how many questions should be DOK 1, DOK 2, and DOK 3. Typically, about 40%-60% of the assessment are DOK 2 questions (Minnesota Department of Education, 2014). I then provide the schools with a bank of math questions that I have collected or created from state test samples and other free resources. Each question is assigned a standard, skill, and DOK. Teachers then use this question bank to create the assessment based on the predetermined number of questions for each standard, skill, and DOK.

### Role and Background of the Researcher

As a coach for schools as they craft these assessments, I see the validity of these questions as essential to the purpose of the test. If these internal assessments are to serve the purpose of indicating a student's level of mastery of the skill, it is essential that the questions are crafted without barriers to the skill. It is incredibly frustrating as a teacher

to create lessons, activities and formative assessments throughout a unit and see great progress only to see students fail to show what they have learned two weeks later on a summative assessment. While students may not have truly mastered the new skills to the level the teacher had thought, it is important to eliminate other possible reasons the student failed on the assessment. When students fail to show mastery of a skill in this model, it is immediately retaught and then reviewed throughout the rest of the school year (Bambrick-Sontoyo, 2010). If a skill is flagged for reteach when it is really an issue of the question, it then takes precedence over skills that may truly need more focus. With further investigation of language patterns that make ELs potentially misunderstand the task, I can include this information in future professional developments and rewrite all necessary questions in the math question bank I share with schools. I can ensure the sample questions I provide avoid more language barriers and coach teachers during professional development to be cognizant of language patterns as they craft and select math word problems for their assessments.

Researchers have already identified the need to remove linguistic complexities from math word problems, but have not extended this research to how consistently it is applied to internal assessments in schools. Therefore, I am studying how consistently the math word problems used on internal assessments follow the guidelines proposed by Universal Design (Abedi & Sato, 2008). Universal Design (UD) is a list of principles that can be used to reduce the linguistic complexities that often make test questions challenging of ELs. I will discuss this further in Chapter 2.

In 2010 I was first tasked to implement DDI at a charter school in the southwest suburbs of Minneapolis. The job of creating math interim assessments for every grade, three times a year seemed overwhelming. I quickly decided that as a school we needed to divide and conquer, so two other lead teachers supported by crafting grades K-2 and grades 3-4 while I crafted assessments for grades 5-8. The task of selecting questions from online samples and our selected curriculum to be more challenging than we thought. We were seeking out questions that were simply the best example of the skills and standards we wanted the students to learn, so we could feel confident that the students had mastered the skill if they could correctly answer the question. The more we understood each skill listed within the MN State Standards, the better we were able to select rigorous yet appropriate questions.

After that first year, it was my task to annually edit these assessments and make them more valid. When examining the assessments across grade level, it became clear that it would be beneficial for one person with an understanding of how each standard grows in complexity from year to year to edit the assessments. For example, the 2<sup>nd</sup> grade assessments were considerably more challenging than the 3<sup>rd</sup> grade assessments. So, in the 2011-2012 school year I redeveloped the assessments to better represent skill progress from grade level to grade level. I also began to get a better grasp on how to determine which standards need to represent more of the assessment. For example, I found that geometry questions in 6<sup>th</sup> grade were only representing about 10% of the assessments, but were representing about 25% of instruction. I worked to make the assessments more valid

by redistributing the number of questions based on the time spent on each standard on the teachers' pacing guides, or curriculum maps.

In the 2012-2013 school year, I hit upon something even more crucial. Even when addressing all the standards and to the appropriate ratio, were we as a school teaching to the appropriate level of rigor? How could we even know? That year Minnesota Department of Education released Test Specifications that clearly define not only the distribution of questions, but distribution of rigor of the questions based on Norman Webb's Depth of Knowledge (DOK). I then edited the assessments for rigor based on these specific levels.

In 2014, I took the position I am currently in, working with multiple schools to edit assessments in the same way. After two years of working to evenly distribute the standards and DOK in each assessment, I am now looking to edit questions based on Universal Design, or selected language modifications, to ensure all questions are equitable for all students. The school where I first began DDI implementation has a very small population of ELs, but the schools I am currently working with are finding it particularly important to ensure students of all WIDA Levels can understand all questions. Word problems about stamp collecting, for example, must be rephrased to things urban immigrant students might collect. I am currently going through school assessments to see if the question word bank I am offering to schools still hold these language barriers, while the Minnesota Department of Education has worked to eliminate language barriers on the Minnesota Comprehensive Assessments (MCAs).

## Research Question

In this study I look to answer the following two questions:

1. *Are the linguistic principles addressed in six distinct areas in Universal design modified in the question bank I provide to craft internal assessments?*
2. *If so, how frequently does this modification occur?*

I will use both qualitative and quantitative measures to collect assessment questions and examine them for the six areas addressed under Universal Design.

## Research and Method

In order to answer these questions, I will examine math word problems from the questions I have collected that are used for grades 3-5 on interim assessments and categorize them by grade, standard, and DOK. I will then examine each question individually and look at the six linguistic areas addressed on standardized tests using Universal Design and determine whether each question follows each area. Universal Design (UD) are the principles used to create language modifications for test questions to ensure equity for all students. These modifications include the following (Abedi, 2008):

- a. Use active voice rather than passive voice.
- b. Avoid negation.
- c. Avoid proper nouns.
- d. Avoid using general language terms that have a special meaning in math contexts.
- e. Reduce written context and be as universal as possible.

- f. To the extent possible, write sentences that are simple and in standard word order.

In chapter 2, I will address limitations of these six modifications while acknowledging that they are a start for equitable test questions. In the data I will have collected, I will look for patterns in the questions violating these principles paying special attention to the standard, grade level, and DOK. I will analyze these patterns to see if we are inaccurately assessing ELs in a specific skill, grade, or DOK and if the questions are violating these principles more often in one area than another.

### Mapping out the Capstone

In the following chapters, I will summarize the relevant research regarding he assessments within the US educational system specifically highlighting assessment of ELS, , address the gap in research, describe the method of my research to address this gap, and finally present my findings and conclusions.

Chapter 2 will specifically give a history of assessment in US schools over the last 125 years, its impact on ELs, the current work to answer this impact, and the relationship between DOK and language development. I will also explain how Universal Design was created, its implications for state testing, and how it can be used to analyze math word problems that may lead to language barriers. Chapter 3 will go into further explanation of the quantitative research paradigm adopted in the current study and the data points that will be used to focus the analysis. I will analyze 60 questions, 20 for grade 3, 20 for grade 4, 20 for grade 5 based on standard, skill, and DOK and determine which of the 6 modifications of Universal Design are violated and how often according to grade, skill, and DOK. Chapter 4 will be devoted to the findings, frequency of principles violated in



different grades, standards, and DOK. Lastly, in Chapter 5 will discuss what patterns of language stood out the most and what type of analysis process I might be able to incorporate into future professional development training and into editing word problems to eliminate language that may make it challenging for students to show what they have learned.

## CHAPTER TWO: LITERATURE REVIEW

The purpose of this study was to perform a text analysis to find out whether math word problems on school-made assessments follow the same linguistic modifications of word problems on standardized tests. By conducting a text analysis, I wanted to answer the following questions:

1. *Are the linguistic principles addressed in six distinct areas in Universal design modified in the question bank I provide to craft internal assessments?*
2. *If so, how frequently does this modification occur?*

With the increased dependence on assessment results to guide instruction in the last ten years, it is crucial to have valid and equitable assessments. Writers of standardized assessments have noticed and worked to address linguistic and cultural barriers. In this chapter, I will begin with a brief history of assessment in the US and the challenges the shift towards increased standardized assessment has had on ELs, specifically in the area of mathematics. I will then speak to the recent modifications assessment writers have made to address these challenges through Universal Design (UD). I will continue by introducing the theory behind Data Driven Instruction (DDI) and schools' implementation of internal assessment. I will end by addressing the gap in current research and how my study may help fill in this gap.

### A Brief History of Assessment

The US educational landscape was not always as it is today. Standardized testing was quite rare in elementary and secondary schools until mid-1900's. In matter of fact,

for the majority of US history education was decentralized and decision making on all areas was left almost exclusively to individual states and academic achievement was not measured across the state let alone the country. It was not until 1954 when *Brown vs. Board of Education* illuminated the inequity of schools that the federal government felt the need to support states in defining what should be taught in schools and hold states accountable to do so (McGuinn, 2006). Equity of education offered to all schools has been at the heart of assessment in US education since it was introduced. Therefore, I believe it to be essential to continually ask if our assessments are staying true to this intention. In this section I will examine the role assessment has taken in education since 1965 in order to give background to the importance of valid, equitable assessments within the school setting.

In 1965 Lyndon B. Johnson took the first pivotal step to incorporating assessment into school life by passing the Elementary and Secondary Education Act (ESEA). This was a response to the fifteen years of reports and outcry to the poor condition of education in the poorest parts of the country (McGuinn, 2006). While no standardized assessments were part of this legislation, federal money was allocated to educational institutions meeting the needs of high poverty students and all funding was denied any school that segregated students based on race. With the goal to increase the quality of education for all students and no measurable outcome, a huge variety of social and educational programs were implemented with these Title 1 funds. It was impossible to measure if these funds were making the difference intended, and the definition of who qualified for these funds was consistently contested (McGuinn, 2006).

During this time, assessments focused on intelligence and cognitive ability, not set curriculum as it is today (Pellegrino, 1999). It was not until 1983 that the federal government started to better define both what academic proficiency meant and which programs were best meeting this.

The search for increased rigor in the classroom began in 1983 when the US Board of Education published *A Nation at Risk*, a report highlighting the issues in the educational system directly related to academic achievement (Gardner, D. P., Larsen, Y. W., & Baker, W., 1983). Within the recommendations, the report appealed to states to adopt more rigorous and measurable standards. Throughout the 1980s and 1990s several initiatives lead to educational goals or standards for every grade level in math and reading and in 2001, George W. Bush with bipartisan congressional support passed the *No Child Left Behind Act of 2001* (McGuinn, No Child Left Behind and the Transformation of Federal Education Policy, 1965-2005, 2006). At the same time, assessments moved away from intelligence and cognitive testing to curriculum and socio-political context testing. In other words, standardized assessments tested a student's set of knowledge and skills taught in the classroom rather than his or her IQ (Pellegrino, 1999). All states then created standards for math and reading and assessed students to show evidence of mastery.

Specific skills within math and reading were highlighted with significant need for further academic support and in 2008 the US Board of Education released *A Nation Accountable*, an examination of the response and improvements made since 1983 (Education U. D.). The report found that classes still were not teaching to the depth and

complexity needed for students to be able to apply skills within content areas into university classes and the work place. Classes revealed to give limited information with little application. The next year, 2009, state leaders from 48 states joined to create the Common Core in order to jointly address the need for increased rigor and depth in math and reading (NGA Center for Best Practices & Council of Chief State School Officers, 2010). In addition, assessments were created not only to represent the content, but the level of rigor to which the content is mastered by students. Around the same time US Board of Education released a guide in how to align standards to DOK (Webb, 2002). In this report Webb's Depth of Knowledge Model for Alignment was introduced as the tool to ensure the appropriate level of rigor was maintained on assessment.

Today many states, including Minnesota, released Test Specifications indicating approximately what percent of the state assessments will represent each level of DOK (Education M. D., 2014). The clear intention is that classrooms will adopt this level of rigor into their instruction, however results have been slow (Boyd, 2008). Studies show teacher made lesson plans, activities, and assessment made little to no improvement in increasing the amount of rigor within the classroom even when state testing expected it. While state standardized tests at minimum consist of 30% DOK 2 and 5% DOK 3 questions, classrooms on average were only at 17% DOK 2 and 2% DOK 3 (Hess, 2009). The result was a drop in test scores and a revealing of need for more fundamental change.

For the last ten years, districts and schools have been asking how to bridge this gap between the rigor of standardized testing and classroom instruction. This question has led many schools to looking at data and finding ways to use the results of assessments

to change their instruction (Larazin, 2014). For example, if most students in seventh grade failed the geometry section of the math state test, schools would look at more curricular support in geometry in seventh grade the following year. Soon many school leaders began to cry out that they were finding out too late that their curriculum and instruction were not serving their students and that they needed to know in time to reteach those same seventh-grade students geometry before they moved to eight grade with this major deficit. This cry was heard by developers of standardized and adaptive tests (such as the NWEA) that tested individual student growth compared to national scores throughout the year and are sold by the private sector. Schools flocked to these with hopes that they would have a better idea of how their students compared nationally to help give an indication of the quality of their instruction (Larazin, 2014). Soon schools found that this data was not enough to change instruction in the classroom. While the assessments show growth against national averages, it does not show the specific skill that students are continually struggling with. Because of copyright laws, teachers are not able to see the specific questions the majority of their students are failing so teachers are often unable to know what students don't know. For example, these tests may reveal that all second grade students are failing in Geometry and Measurement, but teachers still don't know if they are all failing at counting coins or identifying shapes (Bambrick-Sontoyo, 2010).

This continued gap opened the way for Data Driven Instruction (DDI), where schools and districts add internal assessments to their list of testing during the school year. These assessments are created to look just like daily lessons so teachers can know

fairly soon after teaching a lesson if students learned the information sufficiently. They are administered every six to eight weeks to ensure students are retaining what is taught over time, and are cumulative of the information taught over the whole school year (Bambrick-Sontoyo, 2010). When all parts of the DDI are implemented, schools are finally seeing instructional change and students learn with increased understanding and rigor. Assessment is finally doing what is first set out to do.

This, however, leaves schools with an enormous amount of testing each year and reduced instructional time to implement what they have learned through the assessments (Larazin, 2014). Schools in Minnesota on average are testing students in the fall with either a pretest or an adaptive growth test such as the NWEA; internal assessments every quarter to trimester that may take up to three weeks to assess in younger grades; another growth assessment in the winter; the MCA (Minnesota Comprehensive Assessment) in early spring; and finally another growth assessment in late spring/early summer (Minnesota Department of Education, 2014). Teachers are gaining so much data and from so many different sources it becomes challenging to select what data to rely on the most, particularly when the data results from different data sources conflict. On top of it all, teachers have significantly reduced time to teach based on the results, because so much of the school year is spent on testing. One school that I work with added up the number of hours their eighth grade students were testing and found that they spent about one fifth of their year testing. Other schools have reported the same issue. In 2014 a report was released that on average US students are tested twice a month and take up to 20 standardized tests in just one year (Larazin, 2014).

Schools now find themselves wading through all this testing to find a meaningful, valid assessment that can be relied upon so they can reduce the other assessments while using a smaller amount of data to make a real difference in instruction (Larazin, 2014). Approximately one third of the schools I work with have stopped using standardized and nationally normed assessments not required by the state and have become solely dependent on their internal assessments to make instructional growth for the very reasons listed above. This makes the need for these internal assessments to be valid for all students even more urgent so schools can continue with the original purpose of testing, which was to ensure all students, no matter their socio-economic or cultural background, are learning.

To summarize, historically, state standardized testing was established to hold accountability to states and schools to offer rigorous and quality education based on state agreed upon standards to all students (McGuinn, 2006). This took education on a path to the over-tested environment in which schools find themselves today (Larazin, 2014). This path began with the need for states to be held accountable to stop segregation and disproportional educational opportunities based on socio-economic status and race. It started with funding for programs in poor communities in 1965 and became more defined with state standards and standardized testing based on these standards in 1983 with NCLB (McGuinn, 2006). The composition then became more defined in 2008 with the level of rigor to ensure all schools offer all students not only an equitable set of knowledge, but a depth of understanding of that knowledge (Hess, 2009). Schools then found themselves still looking for an assessment that truly changed instruction in the



classroom for all students and began to purchase assessments from the private sector (Larazin, 2014). While these tests showed their students against students nationally, schools had not yet found a test that changed instruction. Through DDI schools have now found that assessment but need to ensure validity of these teacher and school made tests so they can rely on them more and decrease other testing (Bambrick-Sontoyo, 2010).

In the next section, I will discuss specific components of state testing, which schools replicate to ensure validity and a meaningful reflection of state standards on internal assessments. I will then expose the struggle for ELs with a focus on math word problems, how state tests look to answer this, and how this might not be applied in internal school assessments.

#### Current Math Standardized Testing in Minnesota

For this section, I will focus only on Minnesota's math standardized assessments in order to narrow the focus to my question. I will begin with describing the current Math MCA III students take in grades 3-8 and 10. I will describe how they are different than the two previous math MCAs by considering their language modifications, breaking down by standard, skill, and depth of knowledge.

The math MCA III was developed based on Minnesota State Standards that were adopted in 2007. These new standards implement algebra starting in kindergarten and redefined 8<sup>th</sup> grade to the first half of a traditional Algebra 1 course (Education M. D., 2014). Common Core adopted new math standards 2010 and while it did not adopt MN math standards, it used them as a guideline (Common Core, 2013). MN Math Standards

still introduce a few things earlier than Math Common Core. For example, as elapsed time in third grade in the MN Standards (Minnesota Department of Education, 2014).

With a desire for schools to implement state standards to the appropriate focus and rigor, Minnesota Department of Education (MDE) started releasing Test Specifications on an annual basis, making the assessment more transparent and easier to replicate in curriculum design and assessment (Minnesota Department of Education, 2014). The current Test Specifications for the Math MCA III lay out the approximate number of questions each standard receives and the minimum percent of questions at each DOK level. Table 2 and 3 show the current information based on Test Specifications edited November 2015 (Minnesota Department of Education, 2014).

**Table 2: Range of Items per Strand for Mathematics MCA-III Grade 3-8**

Grade	Number of Operational Items	Number & Operation	Algebra	Geometry & Measurement	Data Analysis & Probability
3	42	18–20	6–8	10–13	6–7
4	42	16–20	7–8	10–14	6–7
5	42	15–21	9–13	8–10	6–7
6	42	11–19	10–13	8–11	6–8
7	42	12–16	13–18	7–9	7–9
8	42	6–8	18–29	6–8	6–7

**Table 3: Cognitive Level Target Minimum Distribution**

Grades	Level 1	Level 2	Level 3
3–8	20%	30%	5%

The Test Specifications do not end there. They continue to list each standard and specifically what students will be required to do for each standard in each grade. Under

each standard there is also a list of vocabulary allowed on the MCA III. For example, under the 3<sup>rd</sup> grade standard 3.3.3.4, which represents measurement of temperature using a thermometer, the list of allowable vocabulary is as follows: thermometer, temperature, degrees, increase, and decrease. This is listed for each and every standard tested from grades 3 through 8 and also for grade 10 (Minnesota Department of Education, 2014).

MN Test Specifications also define the level of rigor of each DOK level. Where DOK 1 questions focus on basic recall of memorized facts, DOK 2 questions require students to apply those facts to a situation, or in other words, display the use of a skill or understanding of the concept that makes the fact true (Minnesota Department of Education, 2014). Some examples of DOK 2 math tasks would be to estimate, classify, organize, compare data, or solve multi-step word problems. These tasks require students to make decisions as to how to approach the problem, where a Level 1 question only requires that the students follow a memorized process, perform a well-known algorithm, or demonstrate a rote response (Webb, 2002). Table 4 displays an example of two problems, both under the Minnesota State Standard 3.1.2.1 Addition & Subtraction.

**Table 4: Levels 1 and 2 Depth of Knowledge Standard 3.1.2.1**

Level 1	Level 2
Add.  $\begin{array}{r} 3,954 \\ + 2,646 \\ \hline \end{array}$	What is the missing number?  $682 - 103 = \underline{\hspace{2cm}} - 3$

There are a few unknowns left, however. One big unknown is the number of word problems the Math MCA III contains. There are some standards that require it, such as the 3<sup>rd</sup> grade standard 3.1.2.4 whose description is, “Solve real-world and mathematical problems involving multiplication and division, including both "how many in each group" and "how many groups" division problems.” (Minnesota Department of Education p. 14, 2014) There are others that are not so specific and are assessed both within a word problem and without.

MDE has not released other pertinent information such as the context in which the word problems are asked. There is a list of math specific vocabulary, but whether word problems are going to be asked within the context of a baseball game, a trip on a map, or a class visit to the museum is unknown. Test writers therefore have a set of experiences they assume all students are familiar with, but have not shared what these experiences are so that teachers can ensure their students have the appropriate background knowledge in addition to the vocabulary.

In summary, MDE has taken initiative to create a clear, transparent assessment that can help schools better define what to focus on in instruction of each standard and what vocabulary to eliminate, and what vocabulary to focus on. Schools can also use these guidelines to develop curriculum guides and internal assessments. Internal assessments that reflect these specifications can increase validity and meaningfulness in internal assessments and diminish the need for further, expensive standardized assessments from the private sector. However, schools still need to determine for

themselves the frequency of word problems in some standards and the contexts in which these real-world word problems take place.

### Challenges in Assessing ELs in Math

It is well established that ELs have consistently struggled to show mastery on math assessments even though math is a universal language and therefore should be accessible to all students (Heinze, 2006). So the question arises, are we teaching language development or math? Or both? Are ELs able to understand what is being taught? Are assessments testing language development rather than what students have learned? These questions have been brought up, researched, and discussed at length (Abedi, 2008). In this section I will address the specific challenges that have been faced in math word problems and the linguistic patterns that require advanced language development to understand. I will present these problems in three categories: vocabulary, syntax, and ellipsis and cohesion. I will also pose the potential problems that these barriers to math create in equitable, valid assessment that provides/or offers quality, rigorous education for all students.

Vocabulary has consistently created a barrier for ELs in math word problems. One main reason is the multiple words potentially used to describe one task. Students must know that all of the following could possibly mean addition: *add, plus, combine, and, sum, increased by*. Subtraction, multiplication, division, and equals all have multiple ways they are worded that students must understand to perform the correct task. Even when students understand these cues, they are rarely stated in word problems, expecting

reading comprehension of phrases such as, “how many more,” or “how many in all” for students to know how to answer the word problem (Heinze, 2006).

Another vocabulary barrier is the number of vocabulary words which have specific math meanings that are completely different than those occurring in other familiar contexts. Many of these words are in geometry. For example, a *right* angle or triangle, or a *reflection* of a shape. There are homonyms to math specific words such as *pi* vs. *pie* and *sum* vs. *some* (Heinze, 2006). Other math vocabulary is specific only to math and must be learned alongside these mathematical concepts, making math class and language and math class at the same time.

While modifications cannot address all of these barriers, MDE has worked to address some of these by listing the specific mathematic words students need to know for each standard, limiting the vocabulary ELs need to learn.

Syntax of math word problems also may create linguistic challenges for the EL. Some examples are comparatives, modals, and passive voice. Comparatives using *less than* or *greater than* force the student to determine what the question is asking him or her to do with the two sets of data. Other comparatives, such as *-er*, function the same (Heinze, 2006). For example:

*John is 5 years younger than his brother. If John is seven, how old is his brother?*

Before doing the task, ELs must first examine how the comparative describes the relationship between the two brothers.

Modal auxiliaries are often found in the last sentence of a math word problem, the sentence that indicates what a student is expected to do with the information above (Heinze, 2006). Consider the following example:

*Anne went to the candy store and bought three pieces of candy for \$1.50 each.  
She gave the store worker a five-dollar bill. How much change **should** Anne get back?*

In this example, the student must first consider the following information: three pieces of candy, \$1.50 for one candy, the five-dollar bill. Then, the student must comprehend the modal *should* to understand that he or she must multiply \$1.50 by 3 then subtract the answer from 5

Passive voice creates a similar problem (Heinze, 2006). Students must identify who or what is receiving the action as well as do the math task at hand. For example:

*What number follows 12? (active voice) vs. What number **is followed by** 12?  
(passive voice)*

Here the student must decide if 12 is following the number or the number is following 12 before showing if they understand number order. These are just a few of the syntax issues that make it challenging for ELs to show if they know how to do the mathematical task at hand, but display the struggle word order can create.

Lastly, writers of math problems have a tendency to reduce the repetition of a word problem by leaving out a phrase (ellipsis) that the native speaker can very easily figure out, or by using pronouns (Heinze, 2006). For example:

*Ellipsis:* Mary earns \$15 per hour at her job. John earns four times as much as Mary (*earns* omitted). How much does John earn?

*Pronoun:* Susan had 14 marbles. *She* gave 7 of *them* away. How many does *she* have now?

Both leave the EL to dismantle the language to find out who did what before answering the math word problem. Again, in the next section I will address how MDE has used Universal Design (UD) to modify questions to remove ellipsis and pronouns (Abedi, 2008).

Vocabulary, syntax, and the use of ellipses and pronouns to reduce repetition are a few examples of how math word problems can be more about learning English than about a mathematical equation. While this level of complexity exists in math word problems, it is impossible to measure whether or not a student understands the math standard being assessed. Many of these have been eliminated and reduced on the MCA III through UD, but whether these patterns still exist in internal assessments schools are self-creating is a serious determiner of whether these are valid, meaningful assessments for the EL. In the next section I will connect these language challenges to the six modifications of the MCA III.



### What is Universal Design?

Until 2012, Minnesota had a modified math MCA specifically for ELs and until 2014 a modified math MCA for some students who qualified for Special Education. However, the methods for qualifying students for these modified assessments was complex (Thompson & Thurlow, 2002). The state decided to instead create a universal assessment that took into account language barriers for all students.

Universal design was first adapted from a concept used in architecture to assessment by the National Center on Educational Outcomes (NCEO) in 2002 (Thompson, 2002). The idea is rather than creating multiple different assessments for different learners, to write one assessment that takes all students' needs into consideration. There are seven main elements to universal design, but only Element 6 focuses on readability and comprehensibility (Thompson, 2002). The other elements discuss time given, clear directions, equity of the assessment, environment the assessment is give and legibility of text.

In this paper capstone, I will only discuss Element #6 as it deals specifically with language modifications. Element #6: Readability and Comprehensibility takes several researched paradigms into account (Rokow & Gee, 1987) (Brown, 1999) (Hanson, 1997) and from this research MDE selected the following six principles to attempt to remove language barriers on math assessments:

1. Use active voice rather than passive voice.
2. Avoid negation.

3. Avoid proper nouns.
4. Avoid using general language terms that have a special meaning in math contexts.
5. Reduce written context and be as universal as possible.
6. To the extent possible, write sentences that are simple and in standard word order.

In this section I will break down the reasoning for each of the principles, provide a non-example (example without the modification), example, and discuss other researchers' results on how effective UD is on reducing linguistic barriers in math.

Modification 1: Use active voice rather than passive voice.

Active voice, where the subject is doing the acting, has been consistently found easier to comprehend for ELs than passive voice (Heinze, 2007). Passive voice is when the subject is receiving the action rather than doing the action. The following non-example and example show a problem from standard 3.1.1.5: Compare and order numbers up to 100,000 (Minnesota Department of Education, 2014). The non-example contains the passive voice and the example shows how the problem maintains complexity but removes the passive voice:

*Non-example: What number **is followed by** 12,099? (passive voice)*

*Example: What number follows 12,099? (active voice)*

As mentioned previously, switching between active and passive voice may create confusion as to who or what is receiving the action. By removing passive voice, the test returns the focus to the mathematical calculation (Heinze, 2007).

### Modification 2: Avoid Negation

Negation refers to negating negated verb by placing the word *not* after an auxiliary or modal. There are other ways of forming negation, e.g., by using a prefix, as in appropriate vs. inappropriate. These are not within scope in the current paper. The following non-example and example show a problem from standard 3.4.1.1: Collect, display and interpret data using frequency tables, bar graphs, picture graphs and number line plots having a variety of scales. Use appropriate titles, labels and units (Minnesota Department of Education, 2014).

*Negation: Based on the bar graph, how many students **did not** choose pizza?*

*Avoid Negation: Based on the bar graph, how many students chose noodles, ice cream, and cake?*

Similarl to passive voice, negation can create confusion in what the task of the question truly is (Heinze, 2007).

### Modification 3: Avoid Proper Nouns

Proper nouns, nouns naming a specific person, place thing, or idea, are replaced with common nouns or if a person, a common first name. The following non-example and example show a problem from standard 4.3.1.2: Describe, classify and draw quadrilaterals, including squares, rectangles, trapezoids, rhombuses, parallelograms and kites. Recognize quadrilaterals in various contexts (Minnesota Department of Education, 2014).

*Proper Noun Used: Mrs. Daugherty drew a shape on the board. It has 4 equal sides and no right angles. What shape did she draw?*

*Proper Noun Avoided: The math teacher drew a shape on the board. It has 4 equal sides and no right angles. What shape did she draw?*

While it is not specified in the MDE Math Test Specifications (2014), it is important to note that unfamiliar proper names should be avoided where possible. For instance, Brown (1998) suggests avoiding unfamiliar proper nouns and replacing them with common nouns or first names, not with pronouns. Over- use of pronouns can also create confusion as a student tries to follow which pronoun applies to which previous noun (Gaster & Clark, 1995). This was not included in the modifications of the assessments, but is present in NCEO's report on applying Universal Design to Standardized Assessments (2002).

Modification 4: Avoid using general language terms that have a special meaning in math contexts.

As mentioned earlier, some vocabulary items have specific math meanings that are completely different than those in other familiar contexts. Many of these words are in geometry. For example, a *right* angle or triangle, or a *reflection* of a shape. There are homonyms to math-specific words such as *pi* vs. *pie* and *sum* vs. *some* (Heinze, 2006). MDE has created a list of math-specific vocabulary that they expect all students to know. These words are listed below the specific standard they pertain to on the Test Specifications (Minnesota Department of Education, 2004). Math vocabulary on the questions focus only

on the vocabulary listed below each state standard. For example, the standard for solving real-world multiplication and division word problems in 3<sup>rd</sup> grade specifically states the phrases “how many in each group” and “how many group” will be used as well as specific vocabulary terms multiply, divide, and product.

3.1.2.4 Solve real-world and mathematical problems involving multiplication and division, including both "how many in each group" and "how many groups" division problems.

*Item Specifications*

- Factors are limited to 1–20; 1 factor must have only 1 digit
- Dividend is no greater than 100
- Vocabulary allowed in items: *multiply, divide, product*

(Minnesota Department of Education p.14, 2004).

The 4<sup>th</sup> grade standard for the same skill states looks similar, with a new set of skills and vocabulary but with an additional statement that all vocabulary and skills in previous grade may also be used. This also specifies that vocabulary listed in previous grades will also be included. In this case, there is only one previous standard because the Test Specifications begin in 3<sup>rd</sup> grade. However, this standard combines two separate standards in 3<sup>rd</sup> grade, one for word problems involving addition and subtraction and the other above involving multiplication and division (shown above), and the list on both previous standards must include *operation, strategy and solve*.

Thus, after combining the three standards, the list of math-specific vocabulary that could be potentially used includes: *add, subtract, sum, difference, result, “how many in each group,” “how many groups,” multiply, divide, product, operation, strategy and solve.* The list of vocabulary becomes longer and more difficult to collect with each grade. Without specific training of standards across grade level, teachers can find creating this list very challenging. To add to this, many purchased curricula will have a different list of vocabulary that is incorporated into every lesson and worksheet. The teacher’s task then becomes not only finding MDE’s list, but also eliminating the curriculum’s list of vocabulary that is not on MDE’s list off of each lesson and assignment.

Modification 5: Reduce written context and be as universal as possible.

The last vocabulary modification is to make the written context of real world word problems to be about situations as universal as possible, so students do not have to understand the rules to football, for example, to be able to answer the questions correctly. Since there is not set list of situations listed by the state as “universal” and “as possible” is not measurable, this modification is the most subjective. The following example meets the 4<sup>th</sup> grade standard for real-world math problems involving addition, subtraction, multiplication and division (Minnesota Department of Education, 2004).

*Non-example: A hockey team had \$750 to spend on new equipment. They spent \$543 on new hockey sticks. They spent \$75 on new pucks. How much money does the team have left to spend?*

*Example: A class had \$750 to spend on a classroom library. They spent \$543 on new books. They spent \$75 on new bookshelves. How much money does the class have left to spend?*

It cannot be assumed that all students are familiar with *hockey, hockey sticks, or pucks*. It can be assumed however, that all students are familiar with the classroom setting where it is common to have a classroom library.

Modification 6: To the extent possible, write sentences that are simple and in standard word order.

The last modification is again syntactic. This again reduces the amount of dismantling the sentence to find the mathematical task. The following example meets the expectation for the fourth grade standard 4.1.1.6: Multi-digit Division Strategies (Minnesota Department of Education, 2004).

*Non-example: A group of 245 dogs are going to the dog park and there are 7 vans to take them. If each van has the same number of dogs on board, how many dogs will be in each van?*

*Example: A group of 245 dogs are going to the dog park. There are 7 vans to take them. Each van has the same number of dogs on board. How many dogs will be in each van?*

It is important to note the Universal Design is just one attempt to create an equitable assessment that gives ELs the opportunity to show their understanding of

academic skills. Research is still inconclusive to its rate of success on large scale assessments (Blackorby, Wagner, Cameto, Davies, Levine, Newman, Marder, & Sumi, 2005; Johnston, 2003; Johnstone, Thompson, Moen, Bolt, & Kato, 2005). Sato suggests, “Applying general principles of Universal Design (UD), while necessary, is not sufficient to maximize access vis-à-vis the needs of the ELL population.” (2008, p.58).

Implementing Universal Design in and of itself will not create equitable assessments, but to follow it as the state has done, schools designing their own assessments can come one step closer to honestly assessing ELs understanding of new concepts.

#### School Response to Standardized Testing: Interim Assessments

As mentioned earlier in this section, schools have responded to high-stake state testing first by implementing standardized assessments created and sold by the private sector and secondly by implementing internal assessments into each grade level (Larazin, 2014; Bambrick-Sontoyo, 2010). In this section, my purpose is to first connect the information MDE has given schools about the Math MCA III to the internal, or interim, assessments schools are giving students three to five times a year based on Data Driven Instruction (DDI). I will then discuss the lack of implementation of Universal Design (UD) and how this might affect the meaningfulness of these assessments.

Several school leaders began to use the state standards as the end goal and create school-made assessments that test students on skills quarterly in order to ensure students learned what was taught. With the focus on state testing, it became clear that classrooms can no longer be focused on the instruction, but must change focus to the learning that is taking place (Bambrick-Sontoyo, 2010). Evidence of what the student has learned and



can do after the lesson than the lessons itself becomes the focus. The quality of instruction is determined on what new skills the students can do that they could not do before. The need for a this new focused is highlighted in Data Driven Instruction (Bambrick-Sontoyo, 2010). DDI was then introduced as a method of creating an assessment prior to a six to eight-week instructional period with exactly what standards would be taught during that time frame, and to what depth of knowledge. Then, when that instructional period is over, students would take the assessment with the idea that results would reflect how well they learned what was taught. Teachers plug in each student's answers to each question into a data base to find patterns in incorrect answers. For example, if 75% of students were able to get all math word problems where they must use addition correct but those they must use subtraction wrong. Using this data, teachers analyze which skills or standards students learned most proficiently and which they needed further support. Armed with this knowledge, teachers would reteach and review those standards and then give another assessment in six to eight weeks with the new standards taught as well and the review standards represented. With this readily available data, teachers can break down not only what standard students are struggling to learn, but also zero in on the very basic skill. For example, results from a state assessment will tell a teacher at the end of an academic year that their class failed most questions pertaining to Geometry and Measurement, but not if it was the composition of shapes or the understanding of measurement systems was the root of the problem. With this information, teachers can change both units for the following year, but those students will not benefit from this information (Bambrick-Sontoyo, 2010). With regular or interim

assessments, teachers know directly after the unit what needs to be changed and can address the lack of understanding in the current group of students before moving to another grade level and classroom. This method of assessing, analyzing and taking new action based on results has spread vastly across the US since 2010 (Bambrick-Sontoyo, 2010). Many schools are now training teachers on how to annually edit these assessments and ensure all standards are represented. Taking this further, schools are crafting questions that address the depth of knowledge associated with each question. While teachers learn to write or select assessment questions that measure both recall of information and the application of it to a skill, schools hope to see teachers apply this skill while crafting lesson plans and activities (Bambrick-Sontoyo, 2010).

While schools have not commonly used the six modifications of UD, most schools use questions from state test samples or questions that closely reflect state test samples as models for their questions (Bambrick-Sontoyo, 2010). Therefore, it is unknown as to how many of the questions naturally avoid those vocabulary and syntactical challenges.

In summary, schools are crafting internal interim assessments with the intention of creating tests that reflect the teaching but with an academic expectation comparative to the state's expectation. Teachers and school leaders together are writing assessments six to eight weeks before students take the test and use math state standards, the school curriculum map or pacing guide, and a reflective ratio of DOK questions to create these assessments.

### The Gap in Current Research and Research Questions

Research summarized above suggests that some linguistic aspects may create barriers to measuring ELs true math ability in math word problems, for example use of passive voice may make it harder to determine *who is doing what to whom*.. The summary has also shown that many of these barriers have been modified in standardized assessments that aim to achieve rigor and standard expectations for schools. However, not much research about internal interim assessments has taken place, particularly how well schools implement these modifications to syntax and vocabulary. The current study addresses this gap.

Therefore, I plan to answer to two following questions through my research:

1. *Are the linguistic principles addressed in six distinct areas in Universal design modified in the question bank I provide to craft internal assessments?*
2. *If so, how frequently does this modification occur?*

Through these questions, I plan to learn how well these interim assessments serve teachers in understanding how well their ELs understand math concepts. I will also look for patterns in questions that do not follow Universal Design and look to see if they are concentrated in specific standards and/or levels of DOK.

### Summary

I began this chapter with a brief history of assessment in US schools and expounded on the purpose of accountability through standardized assessment in order to serve all students, no matter their socio-economic status or race. This move to

standardized assessments have filled up instructional hours and offered more data than can realistically be analyzed. Therefore, there is now need for less valid assessments, which make it even more important to create valid internal assessments that truly show academic growth so other assessments can be eliminated.

I then discussed the question of what depth students must understand the information within the standards. Educational experts have come together to attempt to answer this question through leveling tasks into four main areas of rigor (Webb, 2002) and policy makers have since applied these levels to standards and additionally state standardized tests (Minnesota Department of Education, 2014). State test scores began to drop and controversy quickly arose to the validity of assessing these higher tasks and particularly the more complex language these questions tend to use. In order to address this, states implemented Universal Design, a general list of vocabulary and syntactical patterns to avoid on assessments to ensure equity in testing (Abedi, 2008). However, EL students still consistently score significantly below non-ELs in not only reading, but math and science as well.

Laying out the format of DDI and the interim assessments, I explained and how they are crafted based on what is taught and the rigor and expectation of the state. I then went to highlight that while there has been much study on language development and assessment, there has been no focus on whether these principles have played a role in internal interim assessments. I then proposed my two questions to answer this gap. In the next chapter I will discuss my research methods to help schools ensure validity on interim assessments.

### CHAPTER THREE: METHODS

This study was designed to examine the word problems used in interim internal assessments for violations of the six areas of modification in Universal Design. In the analysis the following questions were expected to be answered:

- 1. Are the linguistic principles addressed in six distinct areas in Universal design modified in the question bank I provide to craft internal assessments?*
- 2. If so, how frequently does this modification occur?*

To answer these questions, I will select 60 math word problems from the question bank created to support teachers in the design of math interim assessments. I will use qualitative research methods to identify the UD violations. I will then use quantitative research to find how often each of the six UD modifications are violated and if there is a pattern in the violations. For example, one UD modification is more often violated in DOK 2 questions.

#### Overview of Chapter

In this chapter, I will describe the methodologies used in this study. First, I will provide a rationale and description of the research design and paradigm, a mixed method of both qualitative and quantitative research. Second, I will present the procedures I used to select the study group and collect the data. Next, there will be a description of the 60 math word problems I have chosen from a question bank distributed and used by a group of charter schools. I will then explain the criteria used to select the 60 questions and my

method of analysis of each question. Finally, I will finish the chapter with a conclusion of the methods process and a preview of Chapter Four.

### Research Paradigm

The research paradigm selected for this study contains both qualitative and quantitative aspects. Mixed methods research offers flexibility by combining complimentary research methods into one study (Ivankova & Creswell, 2009). In the last 30 years the practice of combining quantitative and qualitative data has shown to provide a more multidimensional and accurate view of the process of second language acquisition. Mixed methods have been used to collect data in action research, observation, interviews, open response questionnaire items, verbal reports, diaries, and in this case, text (Ivankova & Creswell, 2009). There are four basic mixed methods designs- Explanatory, Exploratory, Triangulation, and Embedded (Ivankova & Creswell, 2009). This study uses the Embedded design.

#### Embedded Mixed Research Paradigm

Embedded Design can be used when a researcher needs to answer a secondary question that requires the use of different types of data (Ivankova & Creswell, 2009). In the case, I must answer the question, *Are the linguistic principles addressed in six distinct areas in Universal design modified in the question bank I provide to craft internal assessments?* This question required qualitative data to be used to identify if and when the six areas of UD are addressed embedded within quantitative data, which indicated how often each of the six areas of UD are addressed. Embedded Design is unique because

it has a predominant method that guides the research study (Creswell et al., 2003). The predominant method in this study is quantitative. Quantitative research is the gathering of numeric data to answer a specific question (Ivankova & Creswell, 2009). In this case, the data I am gathering is the times each of the six UD modifications are violated in each of the 60 selected math word problems. This data cannot be collected however, without using qualitative research. Qualitative research on the other hand, is done by collecting words and images to explore and describe a phenomenon that is being investigated, not to prove or disprove a point (Ivankova & Creswell, 2009). The process of examining the text of math word problems and explaining it through the researcher's understanding of the six UD modifications is the qualitative portion of this study.

This paradigm fits the Embedded Design of mixed methods of research because the main method of research is quantitative. Both research questions look to be answered through the collection of data- the number of occurrences of UD modification violations. This data will be used to identify if questions about a specific standard, grade level, of DOK have a greater tendency to violate the UD modifications. However, this data can only be collected using quantitative research to explore and describe the language within the math word problems.

In conclusion, the advantage of this paradigm for this study is that it allows me to build the study on the well know design of a quantitative text analysis while collecting two types of data (Ivankova & Creswell, 2009). The first type of data is qualitative, observations of the language used in the questions, and the second is quantitative, the

number of times each is of the observed language patterns is used as well as in which categories of questions.

### Data Collection

#### Participants

There are no human participants in this study. Rather, the data is collected from text analysis of word problems. Therefore, no student data or human information is used in this study.

#### Setting

The group of schools that received the question bank vary from urban to suburban schools and vary from small rural schools to medium-sized intercity schools. Similarly, the schools' populations are diverse. Four of the schools are at least 30% EL and at least 85% Free and Reduced Lunch. The others hold a smaller percent of EL, but only one does not have any enrolled. Table 5 shows each school's general location, population, and percent of EL students.

**Table 5 : Schools with Access to the Question Item Word Bank**

<b>Location</b>	<b>Population</b>	<b>Percent of EL</b>
Rural	547	0.5%
Rural	43	0%
Rural	546	10.6%
Suburban	438	2.5%
Suburban	478	9.6%
Suburban	433	1.8%



Suburban	301	3.3%
Urban	193	62.7%
Urban	135	41.5%
Urban	143	34.3%
Urban	552	81.7%

### Materials

Over the course of seven years, I have done extensive study of the MN Math Standards and used this knowledge and information to create and collect math word problems to put into a bank for schools to select from as they create math internal assessments. The questions are from released state items samplers or self-created to match a specific standard when not enough samples could be found. The questions have been used by at least six charter schools in the last 5 years and were edited by a group of teachers and teacher leaders in August 2015. The question bank holds questions for grades Kindergarten through 8<sup>th</sup> grade and are categorized by standard benchmark standard, skill description, DOK, number of points the question should be worth, and the original location of the assessment question. Table 6 shows an example of questions in 4<sup>th</sup> grade.

**Table 6: 4<sup>th</sup> Grade Math Item Sample**

Benchmark	Description	DOK	Points	Question	Answer	Location
-----------	-------------	-----	--------	----------	--------	----------

4.1.1.1	Multiplication and Division Facts	2	1	$\frac{\quad}{+ 4 + 4} \times 4 = 12$ A. 5 B. 6 C. 16 D. 40	A	Created by Lori Magstadt for Math Interim Assessment 4.1
4.1.1.2	Place Value	2	1	What is 15 tens divided by 5? A. 3 ones B. 30 ones C. 30 tens D. 30 hundreds	B	Created by Lori Magstadt for Math Interim Assessment 4.1
4.1.1.3	Multi-Digit Multiplication Procedures	1	1	Multiply. $\begin{array}{r} 95 \\ \times 47 \\ \hline \end{array}$	4,465	Created by Lori Magstadt for Math Interim Assessment 4.1
4.1.1.4	Estimate Products & Quotients	2	1	Estimate the product of 1,398 and 8 by rounding to the nearest thousand. A. 6,000 B. 8,000 C. 9,000 D. 12,000	B	Created by Lori Magstadt for Math Interim Assessment 4.1

There are approximately 3,600 questions from Kindergarten to 8<sup>th</sup> grade, distributed to each standard. The assessments questions vary in DOK to help schools evenly represent the level of rigor of each standard and the original location of the question gives the school a background for the validity of the question.

#### Data Collection Technique 1: Selecting the 60 Math Word Problems

The process of data collection began with selecting questions from the math question item bank. I selected question items using the following steps:

##### Step One

The first step of this analysis was to select a sampling of the thousands of questions I from the question bank described in the last section. I easily narrowed it down to only word problems. To specifically define *word problem*, I selected questions consisting of at least one sentence that students had to comprehend in order to do the problem. Therefore, of a question could be understood based on symbols did not qualify as a *word problem* even if there was a sentence explaining the task of the problem. For example:

*Find the product of the following equation:*

$$95 \times 47 = \underline{\hspace{2cm}}$$

Since the student could determine how to solve the problem using symbols, this question was excluded from the study.

### Step Two

I selected middle grades for my sampling, grades 3, 4 and 5. Third grade is the first grade to take the MCA and the youngest grade on the MCA Test Specifications (Minnesota Department of Education, 2014). These three grades have distinct standards that while different are distinctly related to each other. Consider the three Minnesota geometry standards (2007):

*Grade 3: 3.3.1.1 Identify parallel and perpendicular lines in various contexts, and use them to describe and create geometric shapes, such as **right triangles, rectangles, parallelograms and trapezoids***

*Grade 4: 4.3.1.1 Describe, classify and sketch **triangles**, including **equilateral, right, obtuse and acute triangles**. Recognize **triangles in various contexts**.*

*Grade 5: 5.3.1.1 Describe and classify **three-dimensional figures** including **cubes, prisms and pyramids** by the **number of edges, faces or vertices** as well as the **types of faces**.*

When considering the task required of the students (underlined), it is clear to see the where the verbs indicate and increased understanding of the content. For example, in third grade the student needs only to *identify* and *describe* but must also *classify and sketch* in fourth grade. The same can be done with the information the students are expected to know (bolded). In third grade the focus is on the foundation of shapes, the lines in which create them. It then moves to the 2-D shapes themselves in fourth grade and then to the 3-D figures the 2-D shapes can create in fifth grade.

The distinct progression of task and information is less clear after fifth grade. In sixth grade rate and ratios become embedded within all standards and then all content refocuses on algebra in seventh and eighth grade (Minnesota Department of Education, 2007). With distinct but related standards, possible patterns of violations of the six UD modifications that follow a standard rather than a grade level have to potential to appear.

### Step Three

Similar to Step 2, I looked for standards that built in complexity from third grade to fifth grade. I selected four standards in each grade that are often assessed within a word problem. I also selected standards that were often assessed at different levels of

complexity. For example, I avoided standards that could not be assessed at all DOK levels like the following Minnesota standard in third grade:

3.1.1.3 Find 10,000 more or 10,000 less than a given five-digit number. Find 1000 more or 1000 less than a given four- or five-digit number. Find 100 more or 100 less than a given four- or five-digit number.

The task the student is required to perform (underlined), *find*, does not lend easily to the rigor required to assess DOK 2 and DOK (Webb, 2002).

#### Step Four

The summer of 2015, I created a process of determining the DOK of a math question item based on works by Webb (1999, 2002, 2006) and Vockley (2009). The process first examines the task the student must do. I then examine the content the student must know and determine how often the student has been exposed to this information. For example, in third grade asking a student to use an array to find the answer to  $8 \times 7 =$  is a DOK 2 but asking the same question in fifth grade is DOK 1. Multiplication is introduced for the first time in third grade so it is expected that the student must visualize (a DOK 2 task) to find the answer to  $8 \times 7$ , especially when first introduced in the year. In fifth grade, however, it is expected that after 2 years of exposure and practice  $8 \times 7$  a student will just state a memorized fact (a DOK 1 task) (Webb 2006). Appendix A shows the flow chart of tasks and content used to determine the DOK of each question.

Before selecting questions, I determined to select DOK questions to reflect MDE's Test Specifications (2014). I decided to select 7 DOK 1, 10 DOK 2, and 3 DOK 3

from each grade level. I made sure that the questions selected for each DOK also reflected the same standards from grade to grade. For example, I selected a DOK 3 from each grade level that pertained to data analysis.

### Step Five

Using the matrix in Table 7, I then searched for questions from the question bank that met the following criteria.

**Table 7: Criteria for Questions Selected**

	<b>Grade 3</b>	<b>Grade 4</b>	<b>Grade 5</b>
<b>Standard 1: Multiplication &amp; Division</b>	DOK 1: 2 DOK 2: 3	DOK 1: 2 DOK 2: 3	DOK 1: 2 DOK 2: 3
<b>Standard 2: Real-World Problems</b>	DOK 1: 2 DOK 2: 2 DOK 3: 1	DOK 1: 2 DOK 2: 2 DOK 3: 1	DOK 1: 2 DOK 2: 2 DOK 3: 1
<b>Standard 3: Perimeter &amp; Area</b>	DOK 1: 2 DOK 2: 2 DOK 3: 1	DOK 1: 2 DOK 2: 2 DOK 3: 1	DOK 1: 2 DOK 2: 2 DOK 3: 1
<b>Standard 4: Data Analysis</b>	DOK 1: 1 DOK 2: 3 DOK 3: 1	DOK 1: 1 DOK 2: 3 DOK 3: 1	DOK 1: 1 DOK 2: 3 DOK 3: 1

### Procedure

After selecting the specific questions, I copy and pasted them into spread sheets according to grade level. On one tab, the questions are aligned to the standard and DOK. On the next tab, the spreadsheets are labeled with the question across the top, tied to a standard, and the six principles are list in the left hand column. I read through each question and examined each for the six UD modification. I scored 1pt. for each principle is does not violate. At the bottom of each column, a percent of how many principles were

not violated is provided. On the left hand side in the cell beside each principle the percent of questions that did not violate that principle shows. On a different tab on the spreadsheet the questions are aligned with DOK and also reveal the percent of principles each question does not violate.

### Pilot Study

For the pilot study I selected four fourth grade questions based on the criteria described in the Data Collection section. Table 8 shows the first tab, where the ten questions are aligned to the standard and DOK.

**Table 8: DOK and Standard Alignment**

DOK			
2	There are 153 students going on a field trip to the State Capitol. Students must be put into groups of no more than 7 students per group. How many groups can be created?  A. 21 B. 22 C. 19 D. 18	4.1.1.6	Multi-Digit Division Strategies
2	Two angles drawn together have the sum of the measure of a straight angle. One of the angles is 112 Degrees. What is the measure of the other angle?  A. 68 degrees B. 172 degrees C. 108 degrees D. 72 degrees	4.3.2.1	Measure Angles

2	<p>Jamari drew a shape on paper. She wrote the 2 sentences below about the shape.</p> <p>It has 4 sides.</p> <p>It has 4 right angles.</p> <p>Which of the following shapes did Jamari draw?</p> <p>A. Rectangle</p> <p>B. Triangle</p> <p>C. Circle</p> <p>D. Square</p>	4.3.1.2	Quadrilaterals
2	<p>Mrs. Matschiner practices on a rectangular shaped tennis court. What is the equation for the area of the court in square yards?</p> <p>A. <math>12\text{yd} \times 7\text{yd} = 84\text{yd}^2</math></p> <p>B. <math>12\text{yd} + 12\text{yd} + 7\text{yd} + 7\text{yd} = 38\text{yd}^2</math></p> <p>C. <math>12\text{yd} + 7\text{yd} = 19\text{yd}^2</math></p> <p>D. <math>12\text{yd} \times 12\text{yd} = 144\text{yd}^2</math></p>	4.3.2.3	Area & Multiplication

I then examined each question to see if they followed each of the six UD modifications. Here is the description of the first question:

**There are 153 students going on a field trip to the State Capitol. Students must be put into groups of no more than 7 students per group. How many groups can be created?**

*Use active voice rather than passive voice.* 1pt: no passive voice

*Avoid negation.* 0pts for the use of “no more than”

*Avoid proper nouns.* 0pts for specifying where the field trip is going when it is not important to the question



*Avoid using general language terms that have a special meaning in math contexts.* Opts for using “per” when not specified within the test specifications as the math vocabulary needed for this standard (Education, 2014)

*Reduce written context and be as universal as possible.* 1pt for use of a school context since all students have school experience

*To the extent possible, write sentences that are simple and in standard word order.* Opts for putting “153 students” as the direct object rather than the subject.

Table 9 shows the documentation of the pilot questions on the second tab of the data base used.

**Table 9: 4<sup>th</sup> Grade Pilot Study Results**

A	E	F	G	H	I
		1	2	3	4
<b>Grade 4</b>	<b>COMBINED PROFICIENCY SCORE:</b>	4.1.1.6 Multi-Digit Division Strategies	4.3.2.1 Measure Angles	4.3.1.2 Quadrilaterals	4.3.2.3 Area & Multiplication
Standard		Open	Open	Open	Open
Type		1	1	1	1
Points		1	1	1	1
Answer		1	1	1	1
<b>Student Name</b>	<b>Combined</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
a. Use active voice rather than passive voice.	15%	1	1	1	1
b. Avoid negation.	11%	0	1	1	1
c. Avoid proper nouns.	4%	0	1	0	0
d. Avoid using general language terms that have a special meaning in math contexts.	7%	0	1	1	0
e. Reduce written context and be as universal as possible.	11%	1	1	1	0
f. To the extent possible, write sentences that are simple and in standard word order.	7%	0	0	1	1
		2	5	5	3
	9%	33%	83%	83%	50%

Through the pilot study I found the importance of documenting my explanation or qualitative research as shown above with the first question. I had originally planned to only use the two excel tabs but found that the explanation of why a point was given or not to each modification important to record. With this anecdotal data I can refer back to this same information for future study or to look for even more specific patterns within each of the six modifications. For example, if the last modification (writing simple sentences

in standard word order) is most often violated by adding a direct object when not necessary.

In the next section I will discuss the method used to analyze each question using the 6 modifications of UD.

### Method of Analysis

I will analyze the 60 selected questions from the larger question bank to answer the following questions:

1. *Are the linguistic principles addressed in six distinct areas in Universal design modified in the question bank I provide to craft internal assessments?*
2. *If so, how frequently does this modification occur?*

In order to answer these questions, I will categorize question items by grade level, standard, and DOK. I will use Webb's method as laid out in his handbook (2006) and the flow chart (Appendix A) I created based on this information to identify the DOK of each question. (see Appendix A). I will then examine each question based on the six principles and look for patterns in all questions at a specific grade level, standard at DOK to see if any violate all or one of the six principles more often than another.

As mentioned earlier, the research method I used was a combination of qualitative and quantitative research. While using a specific procedure to select the questions to analyze and each item again the 6 modifications of UD, it is still my judgment that determined both. This text analysis was also quantitative research because statistics were

collected from data (Mackey, Gass, 2005). I collected the percent of times each of the six modifications of UD were violated in the total 60, per grade level, standard, and DOK.

According to Mackey and Gass' description (2005), I am using associational quantitative research, looking for a relationship between variables, not manipulating one variable as experimental quantitative research might. The purpose is to seek a potential relationship between the principles of Universal Design and math word problems on internal interim assessments. There are several variables available to the current study: six variables for Universal Design (the six principles) and three variable on the math assessment side (grade, standard, DOK). A factorial design was adopted, a design that involves more than one independent variable. In total, I will examine 18 relationships between 2 independent variables each time. Table 10 displays these relationships.

**Table 10: Factorial Design**

	<b>UD Modifications</b>						
<b>Question Items</b>		Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod. 5	Mod. 6
	Grade Level						
	Standard						
	DOK						

With 60 questions (each with 3 variables) I will have 90 separate subjects to examine against 6 principles.

For analysis, I first looked at the percent of questions that did not violate each principle in each grade. I then compared the three grades against each other, to see if one

grade violated the principles more often. I also looked at each of the principles to see if one is violated more often in one grade than the other, or if there is one principle consistently violated in each grade. I did this same analysis by standard and DOK.

### Verification of Data

In order to ensure the results are reliable, I will do the three following with each question during the qualitative portion of the research:

1. Follow the same procedure for each question
2. Make anecdotal notes describing the justification for each point given or not given
3. Have the data reviewed by no less than three peer-examiners

The quantitative portion of the research will be verified using excel formulas to calculate percent of violations of each of the six UD modifications.

### Ethics

No student data or human information is used in this study

### Summary

In this chapter I then discussed the Embedded Design mixed method of qualitative and quantitative methods used to examine the relationship between the six principles described in Universal Design and the questions collected. I then described the data collection process of collecting the 60 sample questions analyzed using the 6 UD Modifications. I went on to explain how I will collect and analyze the data in excel spreadsheets and the reliability of the research process. In Chapter Four I will discuss my

findings found from the analysis of the assessment questions. In Chapter Five I will examine these findings, the limitations of the research, implications based on the research and any possible conclusions.

## CHAPTER FOUR: RESULTS

This study examined 60 math word problems from a math question bank schools use in interim internal assessments. The purpose was to look for violations of the six areas of modification in Universal Design to see if internal assessments schools use follow the same linguistic expectations as standardized assessments in Minnesota (Minnesota Department of Education, 2004). The 60 questions were selected based on grade level, standard, and DOK (Depth of Knowledge). 20 questions were selected from third grade questions, 20 from fourth grade, and 20 from fifth grade. These questions were also selected from common standards across the four grade levels: Multiplication and Division, Real World Math Problems, Perimeter and Area, and Data Analysis. A proportional number of questions were also selected based on DOK. These selected questions were examined based on the six areas of modifications in Universal Design. In this chapter I will discuss the results based on each of the six modifications in Universal design, addressing how often each was violated and examples. I will then examine the results of the number of questions that violated the UD Modifications based on grade, standard, and DOK. Through the collection of these data, I sought to find the answer to the following questions:

- 1. Are the linguistic principles addressed in six distinct areas in Universal design modified in the question bank I provide to craft internal assessments?*
- 2. If so, how frequently does this modification occur?*

## Results of Questions Based on the Six Modifications of Universal Design

At least one of the six modifications of UD was violated in 43% of the questions. The most common violation was the use of proper nouns and the least common was the avoidance of negation and simple sentences. The following subsections describe the violations of each modification.

### Modification 1: Use active voice rather than passive voice.

Active voice, where the subject is doing the acting, has been consistently found easier to comprehend for ELs than passive voice (Heinze, 2007). Passive voice is when the subject is receiving the action rather than doing the action. One of the questions, or 2%, used passive voice rather than active voice. Table 11 displays this violation.

**Table 11: Passive Voice Items**

Grade	Standard	DOK	Question
3	Multiplication & Division	1	What is 45 <i>divided by</i> 9?

Switching between active and passive voice may create confusion as to who or what is receiving the action. By removing passive voice, the test returns the focus to the mathematical calculation (Heinze, 2007).

### Modification 2: Avoid Negation

Negation refers to negating negated verb by placing the word *not* after an auxiliary or modal. None of the questions selected used negation.



### Modification 3: Avoid Proper Nouns

Proper nouns, nouns naming a specific person, place thing, or idea, are replaced with common nouns or if a person, a common first name. Twenty-three of the questions, or 38% used a proper noun. Table 12 shows these violations.

**Table 12: Proper Nouns**

Grade	Standard	DOK	Question
3	Real World Math Word Problems	1	<i>Mr. Peterson</i> bought 24 symbols and 37 maracas. How many more maracas did he buy than symbols?
3	Real World Math Word Problems	2	<i>Jeffery's</i> mom is 55 years old. <i>Jeffrey</i> is 32 years younger than his mother. How old is <i>Jeffrey</i> ?
3	Real World Math Word Problems	1	<i>Bobby</i> has 637 baseball cards in his collection. He buys 129 more cards. How many cards does he have now?
3	Real World Math Word Problems	2	<i>Jeff</i> has 96 chapter books and 34 picture books. He sells 15 chapter books and 18 picture books. How many books does he have left?
3	Real World Math Word Problems	3	<i>Jeff</i> had 1,350 glass beads and 695 clay beads. He sold 138 glass beads and 47 clay beads. How many beads did <i>Jeff</i> have left?
3	Data Analysis	1	<i>Danny</i> tossed 2 nickels 10 times. The results are shown in the tally chart below. Which graph shows these results?
3	Data Analysis	2	The graph below shows how many children in <i>Mrs. Nelson's</i> class are 8, 9, and 10 years old. Which of the following is true?
3	Data Analysis	3	In two months, <i>Betty</i> plans to have five times her current savings. If <i>Betty</i> meets her goal, how many dollars will she have saved?
4	Multiplication & Division	2	<i>Mandy</i> handed out four stickers to each of her friends at her party. If she passed out 32 stickers, how many friends were at her party?
4	Real World Math Word Problems	1	<i>Mrs. Moore's</i> preschool class is going on a field trip. The total cost of the trip is \$112. If there are 8 students in <i>Mrs. Moore's</i> class and each

			student pays an equal amount, how much will each student have to pay?
4	Real World Math Word Problems	2	<i>Jonah</i> has 31 boxes of baseball cards. If each box contains 183 cards, about how many baseball cards does <i>Jonah</i> have in his collection?
4	Real World Math Word Problems	2	There are 20 bicycles and 13 cars in the garage at <i>Shereen's</i> apartment building. How many wheels are there in the garage?
4	Real World Math Word Problems	3	<i>James</i> has 37 baseball cards. <i>Paul</i> has 23 more cards than <i>James</i> and 15 more cards than <i>Ron</i> . How many cards do all three boys have?
4	Area & Perimeter	3	<i>Kira</i> is laying 1-inch square tiles to cover a table top. The table is 24 inches by 18 inches. She lays the tiles in strips of six. How many of these does she need to lay with no gaps or overlaps?
4	Data Analysis	2	The table shows how a stock broker followed the stock prices of a certain set of companies. How much more did <i>Thompson Corporation's</i> stock cost on Wednesday than Thursday?
4	Data Analysis	1	The students in <i>Mrs. Dean's</i> class took a vote of their favorite singer. Each student voted for one of five singers. Here is how the vote came out. How many students in <i>Mrs. Dean's</i> class voted?
5	Multiplication & Division	2	<i>Ms. Xiong</i> wants to buy a pencil for each of her 32 students. The pencils come in boxes of 12. How many boxes does <i>Ms. Xiong</i> need to buy? *Which of the following could you do to solve the problem?
5	Real World Math Word Problems	1	<i>Manny</i> owns 83 sets of basketball cards. Each set has exactly 504 cards. What is the total number of basketball cards <i>Manny</i> owns?
5	Real World Math Word Problems	2	<i>Sally</i> went jogging with <i>Jackson</i> . <i>Sally</i> jogged $\frac{5}{8}$ of a mile before she stopped to rest. <i>Jackson</i> jogged $\frac{1}{4}$ of a mile more than <i>Sally</i> before stopping. How many miles did <i>Jackson</i> jog?
5	Real World Math Word Problems	2	<i>Jan</i> has 500 pieces of paper. She prints as many copies as possible of a 16-page report. How many pieces of paper are left?
5	Real World Math Word Problems	3	<i>Mrs. Hilt</i> baked 7 dozen cookies and sold them for \$4.25 per half-dozen. How much money would <i>Mrs. Hilt</i> make if she sold all of the cookies?

5	Data Analysis	2	<i>Maria</i> recorded the heights of 2 plants for 4 weeks. How much did plant 2 grow from week 1 to week 2?
5	Data Analysis	2	<i>Waymen Elementary School</i> is having a fundraiser. The graph below shows how sales increased over six weeks. Approximately how much did sales increase during the first week?

Each grade level had eight questions with proper nouns. Proper nouns were most often found in Real World Math Word Problems (54% of the questions). 29% of the questions were Data Analysis questions and 8% were Multiplication & Division and 8% were Area & Perimeter. One half of the questions are DOK 2 and one fourth are DOK 1 and one fourth are DOK 3. This is proportional to the number of questions at each DOK.

While it is not specified in the MDE Math Test Specifications (2014), it is important to note that unfamiliar proper names should be avoided where possible. For instance, Brown (1998) suggests avoiding unfamiliar proper nouns and replacing them with common nouns or first names, not with pronouns. Over- use of pronouns can also create confusion as a student tries to follow which pronoun applies to which previous noun (Gaster & Clark, 1995). This was not included in the modifications of the assessments, but is present in NCEO's report on applying Universal Design to Standardized Assessments (2002).

Modification 4: Avoid using general language terms that have a special meaning in math contexts.

Some vocabulary items have specific math meanings that are completely different than those in other familiar contexts. Many of these words are in geometry. For example, a *right* angle or triangle, or a *reflection* of a shape. There are homonyms to math-specific words such as *pi* vs. *pie* and *sum* vs. *some* (Heinze, 2006). MDE has created a list of math-specific vocabulary that they expect all students to know. These words are listed below the specific standard they pertain to on the Test Specifications (Minnesota Department of Education, 2004). Math vocabulary on the questions focus only on the vocabulary listed below each state standard. Two, or 3% of the questions, used math specific vocabulary that was not listed for that grade level. Table 13 shows these questions.

**Table 13: Math Specific Vocabulary**

Grade	Standard	DOK	Question
4	Multiplication & Division	2	When the number 4 is <i>placed</i> in the _____, which number sentence is true?
4	Multiplication & Division	1	$8 \times 7$ can be thought of as $(2 \times 4) \times 7 = 2 \times (4 \times 7) = 2 \times 28$ which equals 28 <i>doubled</i> or ____

Modification 5: Reduce written context and be as universal as possible.

This modification is to make the written context of real world word problems to be about situations as universal as possible, so students do not have to understand the rules to football, for example, to be able to answer the questions correctly. Since there is not set list of situations listed by the state as “universal” and “as possible” is not measurable, this modification is the most subjective. There are nine questions (15%) that were noted to use situations that were not noted as universal situations. Table 14 shows these situations.

**Table 14: Universal Context**

Grade	Standard	DOK	Question
3	Real World Math Word Problems	1	Mr. Peterson bought 24 <i>symbols</i> and 37 <i>maracas</i> . How many more <i>maracas</i> did he buy than <i>symbols</i> ?
3	Real World Math Word Problems	1	Bobby has 637 <i>baseball cards</i> in his collection. He buys 129 more cards. How many cards does he have now?
3	Real World Math Word Problems	3	Jeff had 1,350 <i>glass beads</i> and 695 <i>clay beads</i> . He sold 138 <i>glass beads</i> and 47 <i>clay beads</i> . How many beads did Jeff have left?
4	Real World Math Word Problems	2	Jonah has 31 boxes of <i>baseball cards</i> . If each box contains 183 cards, about how many <i>baseball cards</i> does Jonah have in his collection?
4	Real World Math Word Problems	3	James has 37 <i>baseball cards</i> . Paul has 23 more cards than James and 15 more cards than Ron. How many cards do all three boys have?
4	Data Analysis	2	The table shows how a <i>stock broker</i> followed the stock prices of a certain set of companies. How much more did Thompson Corporation's <i>stock cost</i> on Wednesday than Thursday?
5	Real World Math Word Problems	1	Manny owns 83 sets of <i>basketball cards</i> . Each set has exactly 504 cards. What is the total number of basketball cards Manny owns?
5	Real World Math Word Problems	2	Sally went <i>jogging</i> with Jackson. Sally <i>jogged</i> $\frac{5}{8}$ of a mile before she stopped to rest. Jackson <i>jogged</i> $\frac{1}{4}$ of a mile more than Sally before stopping. How many miles did Jackson <i>jog</i> ?
5	Data Analysis	3	What part of 100 of the source of energy is natural gas according to the bar graph below?

The questions that use context noted as non-universal were evenly split between grade level and DOK. However, all but one were Real World Math Word Problems. Some examples of contexts that were noted as universal were situations universal to all schools, saving or spending money, cars and bicycles in a parking lot, and questions about age.

Modification 6: To the extent possible, write sentences that are simple and in standard word order.

The last modification is again syntactic. This again reduces the amount of dismantling the sentence to find the mathematical task. All questions were already broken into simple sentences and held standard word order. None of the questions violated this modification.

Lastly, of the twenty-four questions that violated the modifications, sixteen violated just one and nine violated two. A summary of the number that violated each is show in Table 15:

**Table 15: Number of Modification Violations**

<b>Modification</b>	<b>Number of Violations</b>
Use Active Voice	1
Avoid Negation	0
Avoid Proper Nouns	15
Avoid Math Specific Vocabulary	2
Universal Context	9
Simple sentences in standard word order	0

#### Results Based on Grade Level, Standard and DOK

After examining the questions based on each individual modification, I looked to see what percent of each grade level questions, standard questions, and DOK questions

violated the modifications more often. The purpose was to look for patterns or trends in the data that may help identify larger issues in the language used in math problems.

Approximately 10 % of each of the grade level's questions (Grades 3, 4, 5) violated the modifications. Therefore, reading level did not appear to be a factor in the occurrences of modification violations.

Real World Word Problems had the largest amount of violations, 20% of all Real World Word Problems violated at least one modification, most often being the use of proper nouns and using a non-universal context. 9% of Data Analysis questions, 8% of Multiplication & Division questions, 2% of Area & Perimeter violated a modification.

The types of Depth of Knowledge (DOK) questions that violated the modifications most often were DOK 1 (basic recall questions) at 13%. 11% of DOK 2 (inferential questions), and 7% of the DOK 3 (critical thinking) questions violated a modification.

It is important to note the Universal Design is just one attempt to create an equitable assessment that gives ELs the opportunity to show their understanding of academic skills. Research is still inconclusive to its rate of success on large scale assessments (Blackorby, Wagner, Cameto, Davies, Levine, Newman, Marder, & Sumi, 2005; Johnston, 2003; Johnstone, Thompson, Moen, Bolt, & Kato, 2005). Sato suggests, "Applying general principles of Universal Design (UD), while necessary, is not sufficient to maximize access vis-à-vis the needs of the ELL population." (2008, p.58). Implementing Universal Design in and of itself will not create equitable assessments, but

to follow it as the state has done, schools designing their own assessments can come one step closer to honestly assessing ELs understanding of new concepts.

### Conclusion

The results of this study showed 43% of the sample 60 math questions from a common math question bank violate at least one of the six modifications of Universal Design (UD). The purpose was to look for violations of the six areas of modification in Universal Design to see if internal assessments schools use follow the same linguistic expectations as standardized assessments in Minnesota (Minnesota Department of Education, 2004). The findings were that two modifications were never violated, two in less than 5% of the questions, one in 15% of the questions, and one in 38% of the questions. The 60 questions were selected based on grade level, standard, and DOK (Depth of Knowledge). 20 questions were selected from third grade questions, 20 from fourth grade, and 20 from fifth grade. Approximately 10% of questions in each of these grades violated the modifications. These questions were also selected from common standards across the four grade levels: Multiplication and Division, Real World Math Problems, Perimeter and Area, and Data Analysis. 8% of Multiplication and Division question violated at least one modification. 20% of Real World Math Word problems, 2% of Perimeter and Area, and 9% of Data Analysis violated at least one UD modification. A proportional number of questions were also selected based on DOK. 13% of DOK 1 questions, 11% of DOK 2 questions, and 7% of DOK 3 questions violated at least one UD modification. These selected questions were examined based on the six areas of



modifications in Universal Design. Through the collection of these data, I sought to find the answer to the following questions:

- 1. Are the linguistic principles addressed in six distinct areas in Universal design modified in the question bank I provide to craft internal assessments?*
- 2. If so, how frequently does this modification occur?*

In this chapter I presented the results of my data collections. In Chapter Five I will discuss how my major findings answered the two questions above, their implications, and suggestions for further research.

## CHAPTER FIVE: CONCLUSIONS

In this study, I attempted to answer the following questions:

1. *Are the linguistic principles addressed in six distinct areas in Universal design modified in the question bank I provide to craft internal assessments?*
2. *If so, how frequently does this modification occur?*

I will discuss the answers to these questions based on the major findings in the analysis of math questions used in internal assessments at a group of charter schools. I will discuss the limitations of this study;, the implications to teachers and learning teams as they develop future internal assessments;, how I will share the finding this study;, and further research opportunities.

### Major Findings

The first question I sought to answer was: *Are the linguistic principles addressed in six distinct areas in Universal design modified in the question bank I provide to craft internal assessments?* Based on my results, the answer is mixed. Four of the distinct areas of modification are consistently addressed, while two less often. Passive voice was only used once, negations were avoided in all questions, math specific vocabulary not identified by the state was only present in two questions, and all questions were in simple sentences or standard word order. The questions however, did not as often address the avoidance of proper nouns or the use of universal context for real world word problems. While proper nouns were used more often (in 1 out of 4 questions), there is contradicting research as to the use of proper nouns and whether they help or hinder comprehension for

ELLs (Blackorby, Wagner, Cameto, Davies, Levine, Newman, Marder, & Sumi, 2005; Johnston, 2003; Johnstone, Thompson, Moen, Bolt, & Kato, 2005). However, other researches highlight the challenge context unfamiliar to ELLs create in math word problems (Heinze 2006).

The second question I sought to answer was: *If so, how frequently does this modification occur?* In order to answer this, I looked at how often the two modifications that occurred less often could be applied. I also looked to see if there were patterns of violations of all modifications within grade level questions, standard questions, and Depth of Knowledge questions.

The first modification that was violated over 15% of the time was the use of proper nouns. Proper nouns were used in 23 of the 60 questions. The occurrences happened most often in real world math problems. I found that all but three of the questions could easily be modified by changing the proper noun to a common noun such as “a student,” or “a gardener.” Two of the questions used the possessive form of the proper noun: *Mrs. Nelson’s class* and *Sharren’s apartment building parking lot*. I concluded these two questions were most in need for modification due to the addition of so many words unnecessary to the meaning of the question. Two of the questions relied on the proper nouns to differentiate between two people. Research shows the over- use of pronouns can also be very confusing, so in the following two questions I concluded it would be difficult to avoid the use of proper nouns, even though the first is used in the possessive form in the first sentence (Abedi & Sato, 2008).

Question Item: *Jeffery's mom is 55 years old. Jeffrey is 32 years younger than his mother. How old is Jeffrey?*

Question Item: *Sally went jogging with Jackson. Sally jogged  $\frac{5}{8}$  of a mile before she stopped to rest. Jackson jogged  $\frac{1}{4}$  of a mile more than Sally before stopping. How many miles did Jackson jog?*

The last question I concluded that could not be changed was a question referring specifically to one company on a data analysis questions. However, there are other issues with this question I will address with the next modification, universal context.

Question Item: *The table shows how a stock broker followed the stock prices of a certain set of companies. How much more did Thompson Corporation's stock cost on Wednesday than Thursday?*

Overall, I concluded that while the use of proper nouns was avoided in the majority of questions, it could easily be avoided more often without adding confusion by replacing these proper nouns with pronouns.

The second modification that was violated more than 10% of the time was the use of context that was not universal. Three of the nine questions could be modified easily by changing the items in the question to more universal items. For example, a question has a music teacher buying symbols and maracas. It would be simple for a teacher to change these items to items used in the music class in at their school. Six questions, however, would be much more challenging to modify and may need to be replaced completely. Three of these questions were about collecting baseball cards and on collecting basketball

cards. Since children do not collect items such as sports cards as a hobby like they did 20 years ago, I found it challenging to simply replace baseball cards with a different item that children might collect. The last two were the data analysis questions: the first was about the stock market and the second about energy sources. The stock market and energy sources are central to these questions and could not simply be renamed something more universal to a student's life. Therefore, I concluded that teachers will need training and support to be able to identify universal settings so they can use this knowledge as they select questions since many of the questions do not follow this modification and could not be easily changed to do so.

I then examined the percent of questions that violated at least one modification based on grade level, standard, and depth of knowledge (DOK). There was an equal number of questions that violated each grade level, grade 3, 4 and 5. I concluded that it was significant that the number of violations did not correlate to the reading level of the text since the reading difficulty increases with the grade level. Real World Math Word Problems violated modifications significantly more often (20% of the questions) than the other three standards due to the use of proper nouns and the need for context to create these "real world" situations. Multiplication & Division and Data Analysis violated the modifications about 9% of the time. Both were mostly due to the use of proper nouns. This lead me to conclude that the issue remains with the need of avoiding proper nouns more often and finding more universal contexts for math problems.

The results from the DOK level was more surprising. 13% of DOK 1 questions, 11% of DOK 2 questions, and 7% of DOK 3 questions violated at least one UD

modification. I did not expect to see the simple questions, DOK 1, violate the modifications more often and see the most rigorous questions, DOK 3, violate the modifications less often. To try to find a reason, I examined what types of violations each question made based on DOK. I noticed that DOK 1 questions most often violated the use of proper nouns where DOK 2 and DOK 3 questions used proper nouns less often. Therefore, I conclude that these higher order questions have been more carefully crafted to ensure linguistic equity than the less rigorous questions. However, most of these DOK 1 questions could easily go through the same modifications, as mentioned in the proper noun conclusion earlier.

The main conclusion is that there is clearly a need for more questions with universal context to be made available for teachers. These questions would be much more challenging to modify than the questions that use proper nouns. There are also many questions available that do not use a universal context, so training teachers to identify these so they might avoid selecting them would also be important.

### Limitations

The limitations to this study revolve around two main areas. The first is the need for more modifications than just the six modifications proposed in Universal Design. Research shows that this is just a first step to crafting math questions that reduce the need for language proficiency (). With so many linguistic components that obstruct assessing math skills without the need for language proficiency, this it is challenging to select the right number of components without making it too broad or too specific.

The second limitation is the questions selected. When considering all the questions available to teacher on the internet, 60 is a small number and the conclusions can only be representative of the question bank I have developed, not all questions available to teachers when crafting internal assessments.

### Implications and Further Steps

The major implications of this study include the need for more work in the context of real world problems and data analysis questions. Many of the questions have outdated context or context that is not familiar to middle and upper elementary students. The big question is, other than questions about school, what contexts can be defined as universal? While replacing the questions with non-universal questions on the question bank is essential, I think this will not get to the root of the issue. Teachers will still search other places for math questions and professional development on identifying questions with universal context. More powerful would be professional development on how to craft math world problems based on science and social studies units that are being taught concurrently. Since these questions are internal, I believe the context can be universal to the student body, not necessarily to all students of that grade level like the MCA questions need to be. This would establish universal context since all students are learning about the context and can draw real meaning and use of math in multiple contexts.

Considering the implications discussed above, I will first modify the questions in both the proper noun category and the universal context on the math questions bank and replace those that cannot be modified. This will go directly to teachers utilizing the math

question bank. I will then make a list of common universal contexts to provide teachers during math professional development opportunities I provide through my job.

While not perfect, a better understanding of the six modifications of UD is a place for teachers to start when creating math assessments, whether for a unit or end of the week quiz. Offering professional development to the schools I serve that gives a clear definition of each and a few examples could be powerful in expanding equity in word problems, with a focus on the context of the word problems. While some of the schools I work with are in areas with few ELs (as referenced in Table 5), all schools but one have at least one and therefore could benefit from this training.

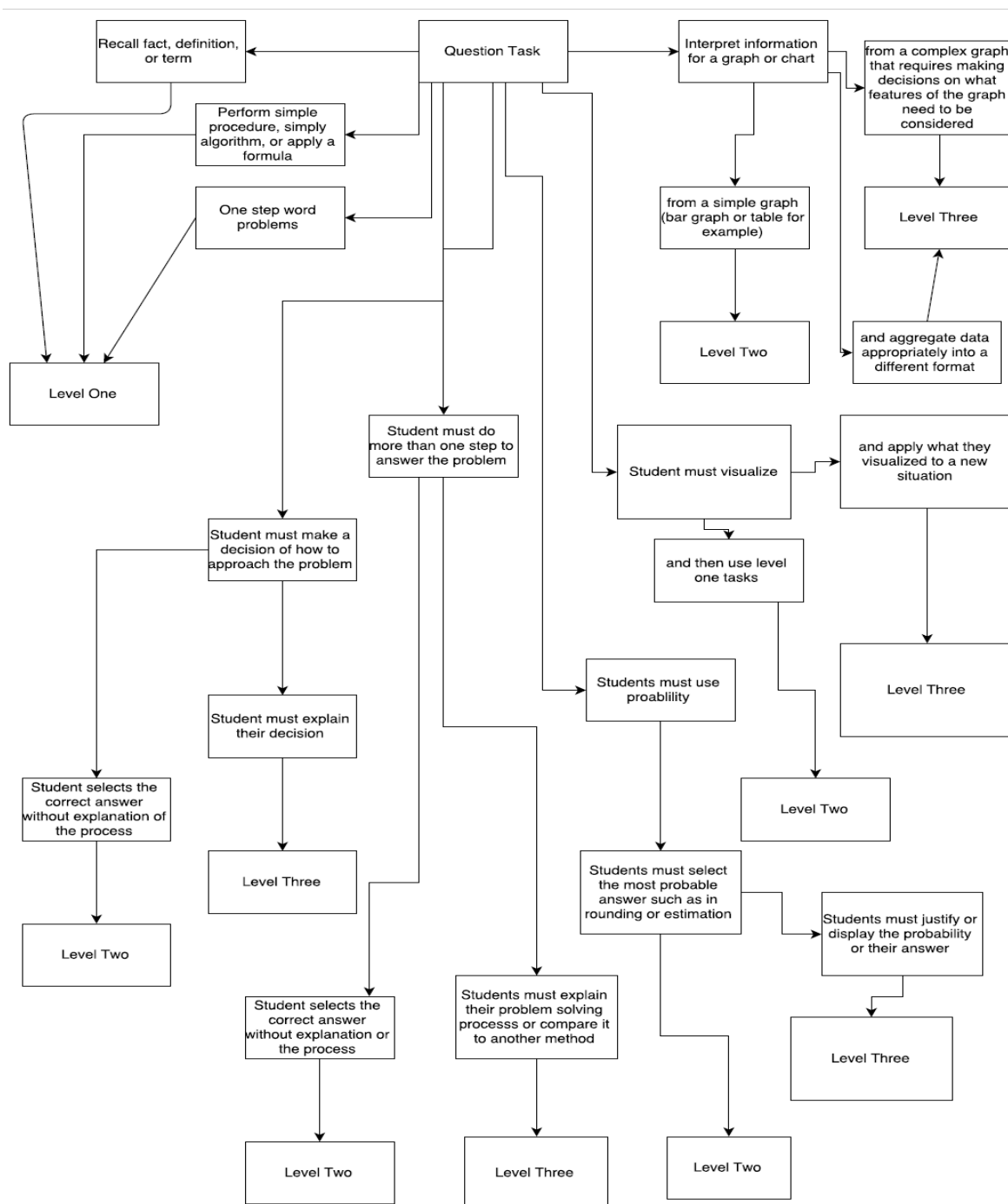
Lastly, I will start to do more research on how to teach writing real world math word problems so I might eventually provide teacher training in this area for teachers to create questions unique to their student body with greater confidence.

#### Further Research

Math word problems in real world contexts drew my attention during this study. The major question this study generated for me was how we can make real world math questions more universal in context. More research on how to embed math into science and social studies is needed so that students see its value and teachers have a larger number of contexts they feel confident their students know well enough to answer math problems about. It needs to be investigated further if such a cross-curricular approach would make learning math concepts easier for ELLs or at least increase motivation to learn how to apply math concepts in real-life.



## APPENDIX A



## REFERENCES

- Abedi, J. & Sato, E. (2008). *Linguistic modification. Part I: Language factors in the assessment of English language learners: The theory and principles underlying the linguistic modification approach; Part II: A guide to linguistic modification: Increasing English language learner access to academic content*. Washington, DC: U.S. Department of Education: LEP Partnership. Retrieved on January 25, 2015, from [http://www.ncela.us/files/rcd/BE024210/Linguistic\\_Modification.pdf](http://www.ncela.us/files/rcd/BE024210/Linguistic_Modification.pdf).
- Bambrick-Sontoyo, P. (2010). *Driven by data: A practical guide to improve instruction*. San Francisco, CA: Jossey-Bass.
- Blackorby, J., Wagner, M., Cameto, R., Davies, E., Levine, P., Newman, L., Marder, C., & Sumi, C. (2005). *Engagement, academics, social adjustment, and independence: The achievements of elementary and middle school students with disabilities*. Paper presented at the Technology and Persons with Disabilities Conference at the Center on Disabilities, California State University, Northridge.
- Boyd, B. T. (2008). *Effects of State Tests on Classroom Test Items in Mathematics. School Science And Mathematics*, 108(6), 251-262.
- Brown, P. (1998). Conversational structure and language acquisition: The role of repetition. *Journal of Linguistic Anthropology*, 8(2), 197-221.
- Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. *Handbook of mixed methods in social and behavioral research*, 209-240.
- Dore, J. (1974). A pragmatic description of early language development. *Journal of psycholinguistic Research*, 3(4), 343-350.
- Gardner, D. P., Larsen, Y. W., & Baker, W. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: US Government Printing Office.
- Gaster, L., & Clark, C. (1995). *A Guide to Providing Alternate Formats*.
- Hess, K., Carlock, D., Jones, B., & Walkup, J. (2009). What exactly do “fewer, clearer, and higher standards” really look like in the classroom? Using a cognitive rigor matrix to analyze curriculum, plan lessons, and implement assessments. Paper

presented at Council of Chief State School Officers (CCSSO), Detroit MI, June 2009. Retrieved from <http://www.nciea.org/cgi-bin>.

Ivankova, N. V., & Creswell, J. W. (2009). Mixed methods. *Qualitative research in applied linguistics: A practical introduction*, 135-161.

Minnesota Department of Education. (2007). *Minnesota K-12 Academic Standards in Mathematics*. Retrieved on January 25, 2015, from <http://education.state.mn.us/MDE/EdExc/StanCurri/K12AcademicStandards/Math/index.html>

Minnesota Department of Education. (2014) *Mathematics Test Specifications for MCA III for Grades 3-8*. Retrieved on January 8, 2015, from <http://education.state.mn.us/MDE/EdExc/Testing/TestSpec/>

Moschkovich, Judit. (2012) *Mathematics, the Common Core, and Language: Recommendations for Mathematics Instruction for ELs Aligned with the Common Core*. University of California, Santa Cruz, CA.

Ockey, Gary J. (2007) *Investigating the Validity of Math Word Problems for English Language Learners with DIF*, *Language Assessment Quarterly*, 4:2, 149-164, DOI: 10.1080/15434300701375717

SciMathMN & Minnesota Department of Education (2015) *Mathematics Frameworks*. Retrieved on January 25, 2015, from <http://education.state.mn.us/MDE/EdExc/StanCurri/K--12AcademicStandards/Math/index.html>

Spellings, M. (2008). *A nation accountable: Twenty five years after a nation at risk*. Retrieved February 25, 2015, from <http://www2.ed.gov/rschstat/research/pubs/accountable/accountable.pdf>.

Vockley, M. (2009). *Alignment and the States: Three Approaches to Aligning the National Assessment of Educational Progress with State Assessments, Other Assessments, and Standards*. The Council of Chief State School Officers. Retrieved January 25, 2010.

Webb, N. L. (1999) *Alignment of science and mathematics standards and assessments in four states (Research Monograph No. 18)*. Madison: University of Wisconsin – Madison, National Institute for Science Education.

Webb, N. (2002, March 28). *Depth-of-Knowledge Levels for Four Content Areas*. Unpublished Paper.

Webb, N. L. (2006). Identifying content for student achievement tests. *Handbook of test development*, 155-180.

Wiggins, G. P., McTighe, J., Kiernan, L. J., & Frost, F. (1998). *Understanding by design* (pp. 0-87120). Alexandria, VA: Association for Supervision and Curriculum Development.